

De briljantie van SPARQL

Dimitri van Hees, PLDN 26 november 2014

Introductie

- Technisch Architect
- Data Specialist
- Freshheads BV
- 'API first' fan





Nederland – Argentinië: 2 - 4



Proof Of Concept

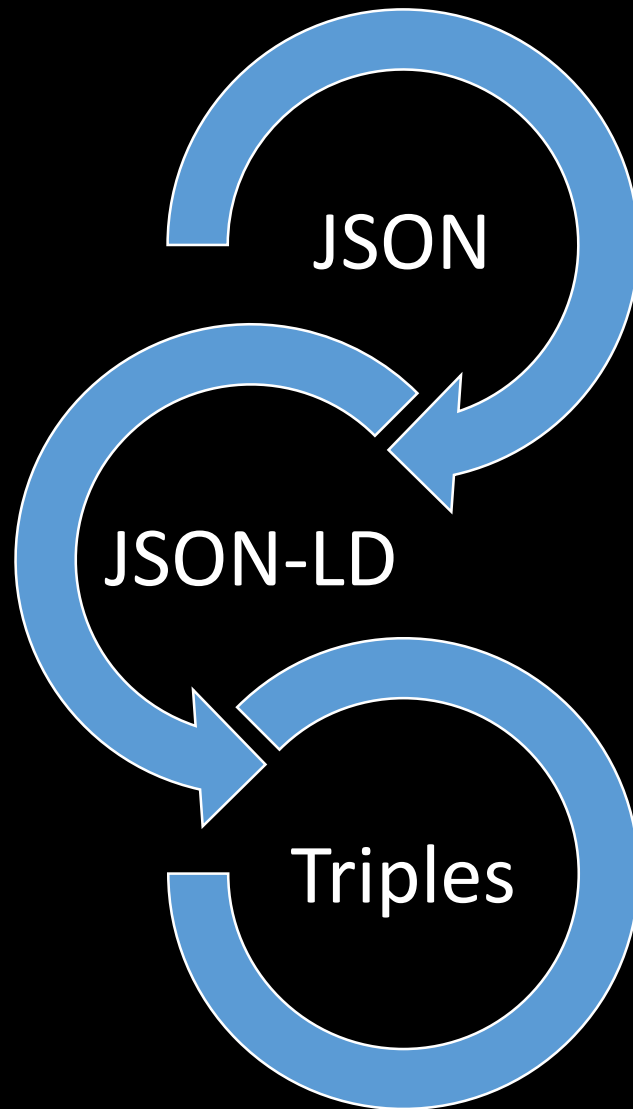
“Build a dashboard with which we are capable to better learn about our fans and customers so that we can take action on that knowledge to win sympathy for our brand”

Beschikbare data

- Facebook (likes, favorites, education, profile)
- LinkedIn (skills, education, jobs, profile)
- Twitter profiles
- Website profiles

Uitdagingen

- Intern
- Extern
- Technisch
- Agile/rapid development



JSON

```
{  
  "id": 43274,  
  "name": "Dimitri van Hees",  
  "gender": "m",  
  "birthDate": "1984-03-14",  
  "birthPlace": "Nijmegen",  
  "facebookLikes": [  
    69116329538,  
    272002449634285,  
    114690771875816  
  ]  
}
```

JSON-LD Context

```
{  
  "id": "@id",  
  "name": "http://xmlns.com/foaf/0.1/name",  
  "gender": "http://xmlns.com/foaf/0.1/gender",  
  "birthDate": "http://schema.org/birthDate",  
  "birthPlace": "http://dbpedia.org/ontology/birthPlace",  
  "facebookLike": "http://xmlns.com/foaf/0.1/interest"  
}
```

JSON-LD

```
{
  "id": "http://example.com/customers/43274",
  "type": "schema:Person",
  "name": "Dimitri van Hees",
  "gender": "m",
  "birthDate": "1984-03-14",
  "birthPlace": {
    "id": "http://dbpedia.org/resource/Nijmegen", "type": "schema:City"
  },
  "facebookLikes": [
    { "id": "http://graph.facebook.com/69116329538", "type": "schema:MusicGroup" },
    { "id": "http://graph.facebook.com/272002449634285", "type": "schema:TVSeries" },
    { "id": "http://graph.facebook.com/114690771875816", "type": "schema:City" }
  ]
}
```

Van gegevens naar informatie



JSON-LD

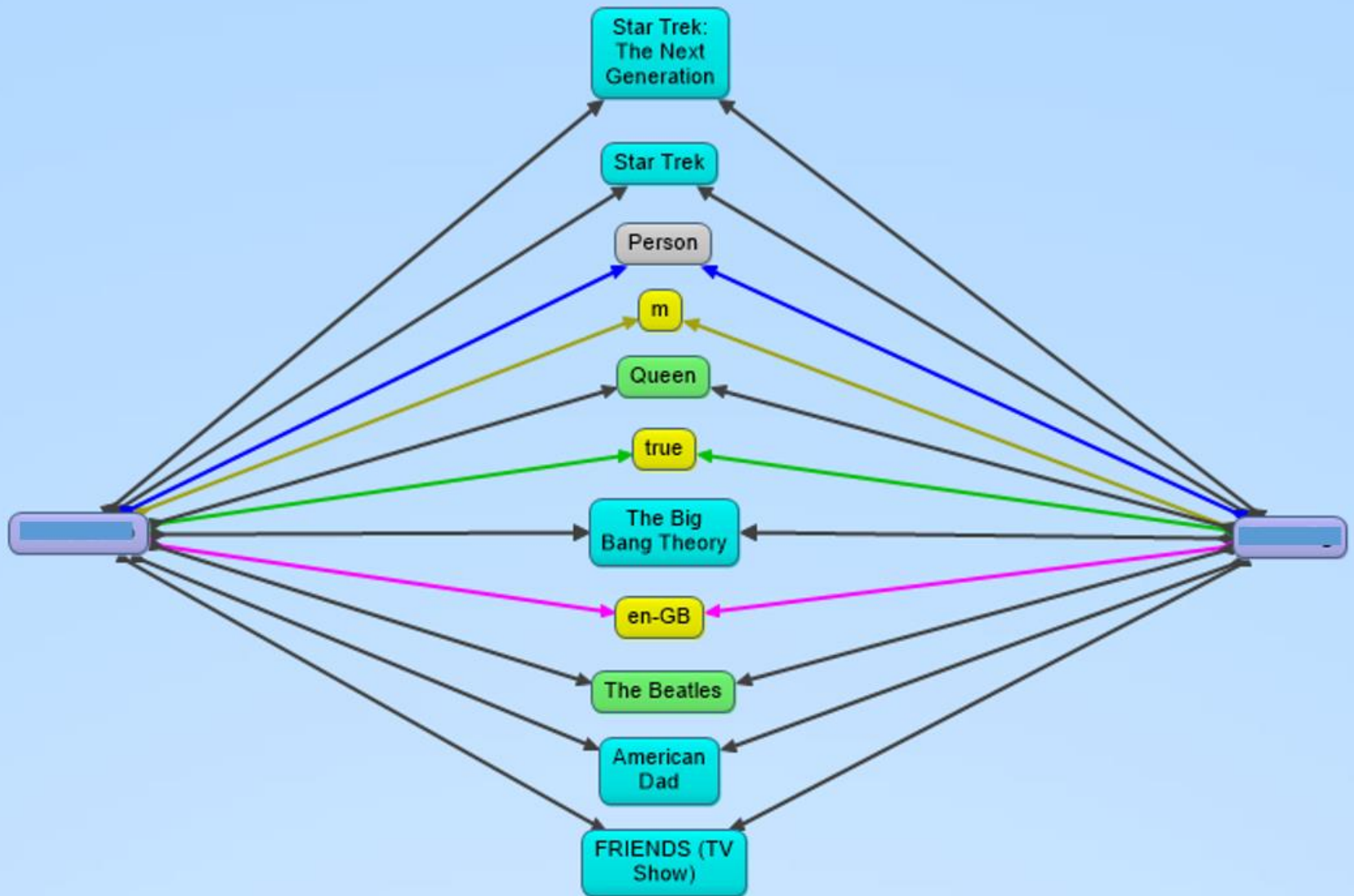
```
{
  "id": "http://example.com/customers/43274",
  "type": "schema:Person",
  "name": "Dimitri van Hees",
  "gender": "m",
  "birthDate": "1984-03-14",
  "birthPlace": {
    "id": "http://dbpedia.org/resource/Nijmegen", "type": "schema:City"
  },
  "facebookLikes": [
    { "id": "http://dbpedia.org/resource/The_Beatles", "type": "schema:MusicGroup" },
    { "id": "http://dbpedia.org/resource/Monty_Python", "type": "schema:TVSeries" },
    { "id": "http://dbpedia.org/resource/Nijmegen", "type": "schema:City" }
  ]
}
```

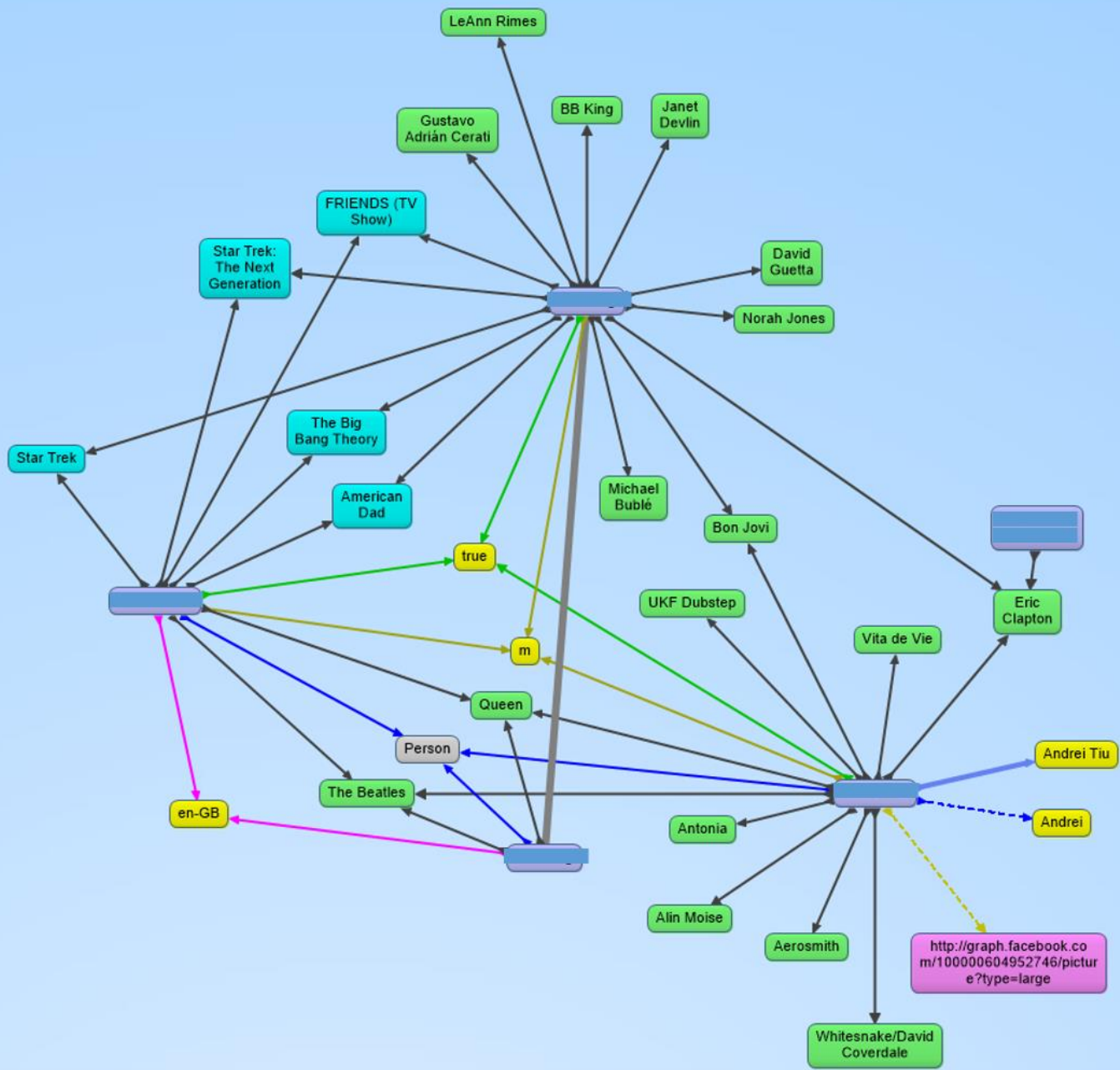
Triples

Subject	Predicate	Object
http://example.com/customers/43274	http://xmlns.com/foaf/0.1/name	"Dimitri van Hees"
http://example.com/customers/43274	http://xmlns.com/foaf/0.1/gender	"m"
http://example.com/customers/43274	http://dbpedia.org/ontology/birthPlace	http://dbpedia.org/resource/Nijmegen
http://example.com/customers/43274	http://xmlns.com/foaf/0.1/interest	http://dbpedia.org/resource/The_Beatles
http://example.com/customers/43274	http://xmlns.com/foaf/0.1/interest	http://dbpedia.org/resource/Monty_Python
http://example.com/customers/43274	http://xmlns.com/foaf/0.1/interest	http://dbpedia.org/resource/Nijmegen

Triplestore: AllegroGraph

- Gratis tot 5.000.000 triples
- SPARQL 1.1 support
- REST API toegang
- Lokale installatie was nodig i.v.m. privacy





SPARQL 1.1: federated queries

- Wie van mijn klanten is ouder dan 40 en woont in een hoofdstad?
- Wie van mijn klanten woont in een stad met meer dan 600.000 inwoners?
- Wie van mijn klanten vindt een bepaalde film leuk van een bepaalde regisseur?
- Etc.

**Wie van mijn klanten
vindt een bepaalde film leuk
van een bepaalde regisseur?**

**Wie van mijn
klanten**



Eigen data

**vindt een bepaalde
film leuk**



Externe data

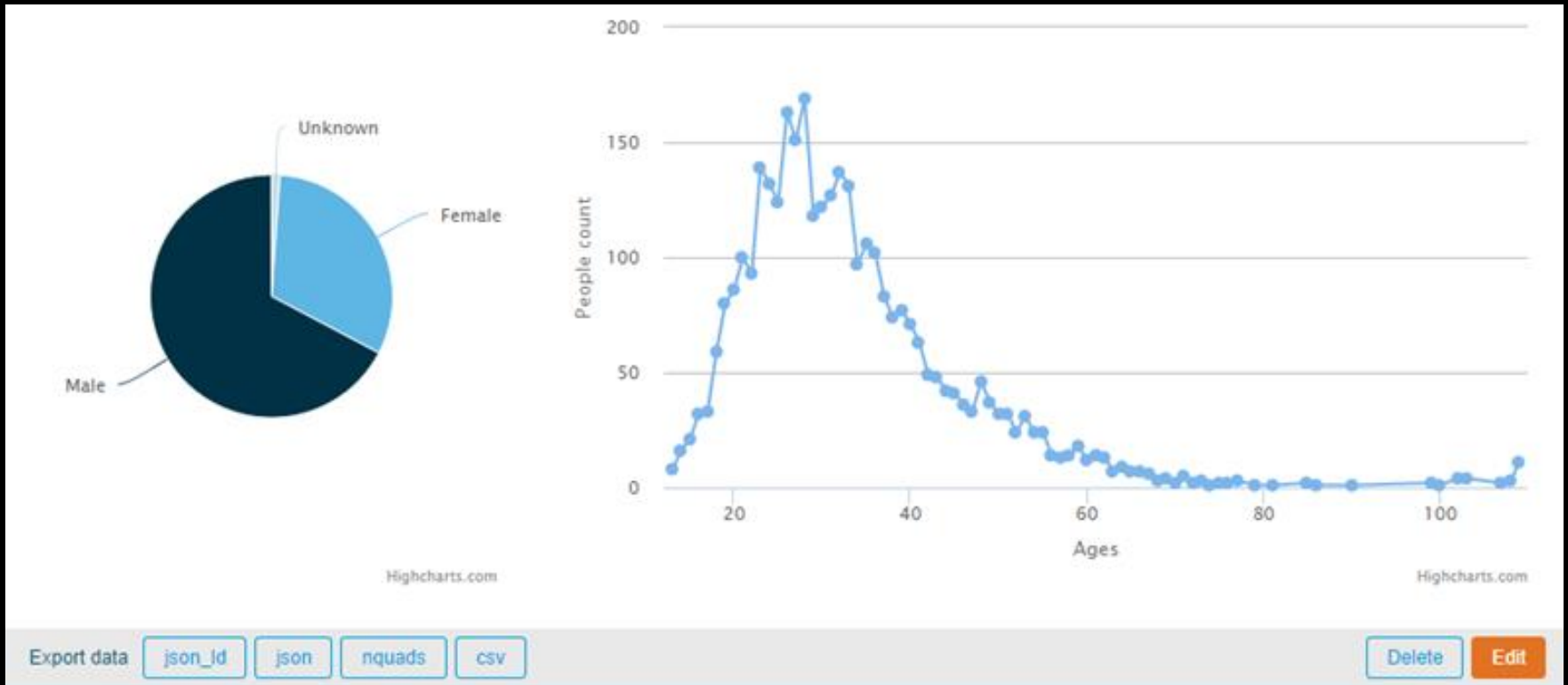
**van een bepaalde
regisseur?**



Open data

...?

Wie van mijn klanten woont in een stad met meer dan 600.000 inwoners?



Wie van mijn klanten woont in een stad met meer dan 600.000 inwoners?

TV Series

	The Simpsons 191 (25.67%)
	The Big Bang Theory 165 (22.18%)
	South Park 157 (21.10%)
	House 155 (20.83%)
	Family Guy 152 (20.43%)

Artists

	The Beatles 744 (100.00%)
	Queen 279 (37.50%)
	Pink Floyd 265 (35.62%)
	The Rolling Stones 228 (30.65%)
	Coldplay 225 (30.24%)

NGO's

	Avaaz 16 (2.15%)
	Breast Cancer Prevalence 16 (2.15%)
	Zonder baard, geen kleding 14 (1.88%)
	(RED) 14 (1.88%)
	AnonymousBrasil 13 (1.75%)

Cities

New York City 61 (8.20%)
Amsterdam 41 (5.51%)
London, United Kingdom 33 (4.44%)
Paris 21 (2.82%)
Los Angeles 16 (2.15%)

Games

	League of Legends 19 (2.55%)
	Need for Speed 14 (1.88%)
	Plants vs. Zombies 11 (1.48%)
	Silent Hill 11 (1.48%)
	Pro Evolution Soccer 11 (1.48%)

Barcelona

Barcelona Swing Vintage

Barcelona

Barcelona vs Real Madrd

Barcelona, Anzoátegui

Barcelona

```
1 PREFIX schema: <http://schema.org/>
2 PREFIX dbpedia: <http://dbpedia.org/ontology/>
3
4 SELECT DISTINCT ?person
5 {
6   ?person a schema:Person .
7   ?person schema:homeLocation ?homeLocation .
8   SERVICE <http://dbpedia.org/sparql> {
9     ?homeLocation dbpedia:populationTotal ?population .
10    FILTER (?population > 600000)
11  }
12 }
```

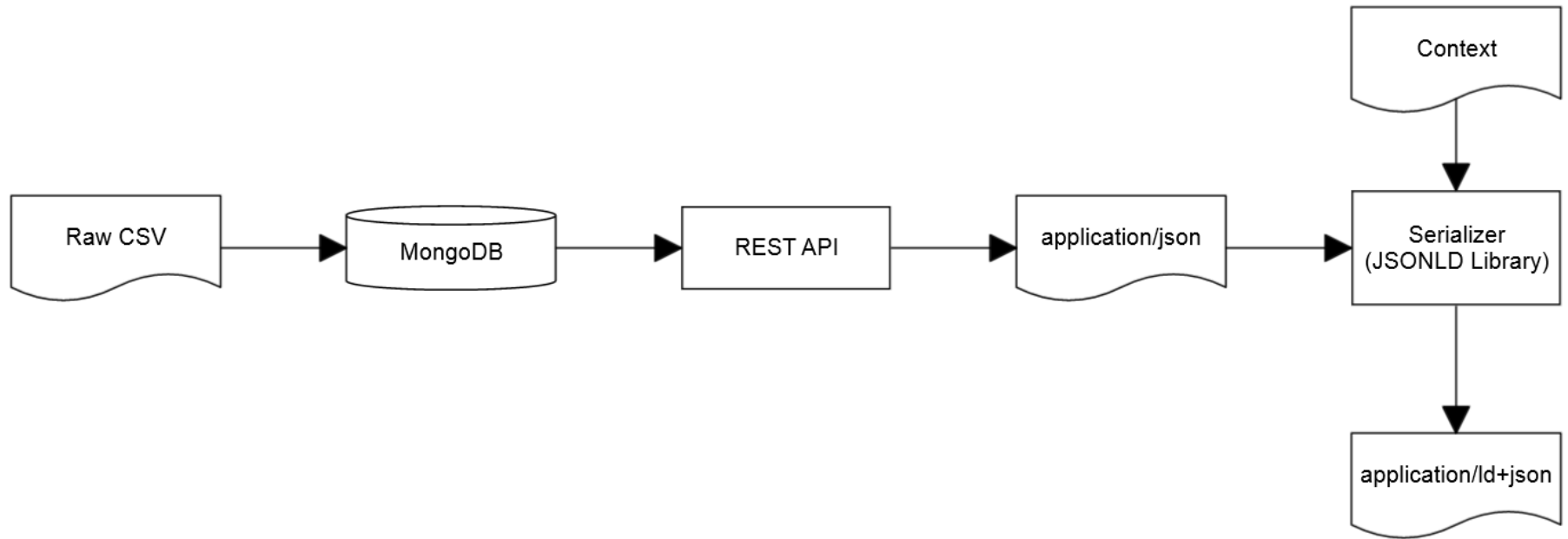
Eén API, drie functies

- Conversie naar Linked Data
- Export functie met behulp van content-negotiation: JSON, JSON-LD, CSV, Triples
- Applicatie interface

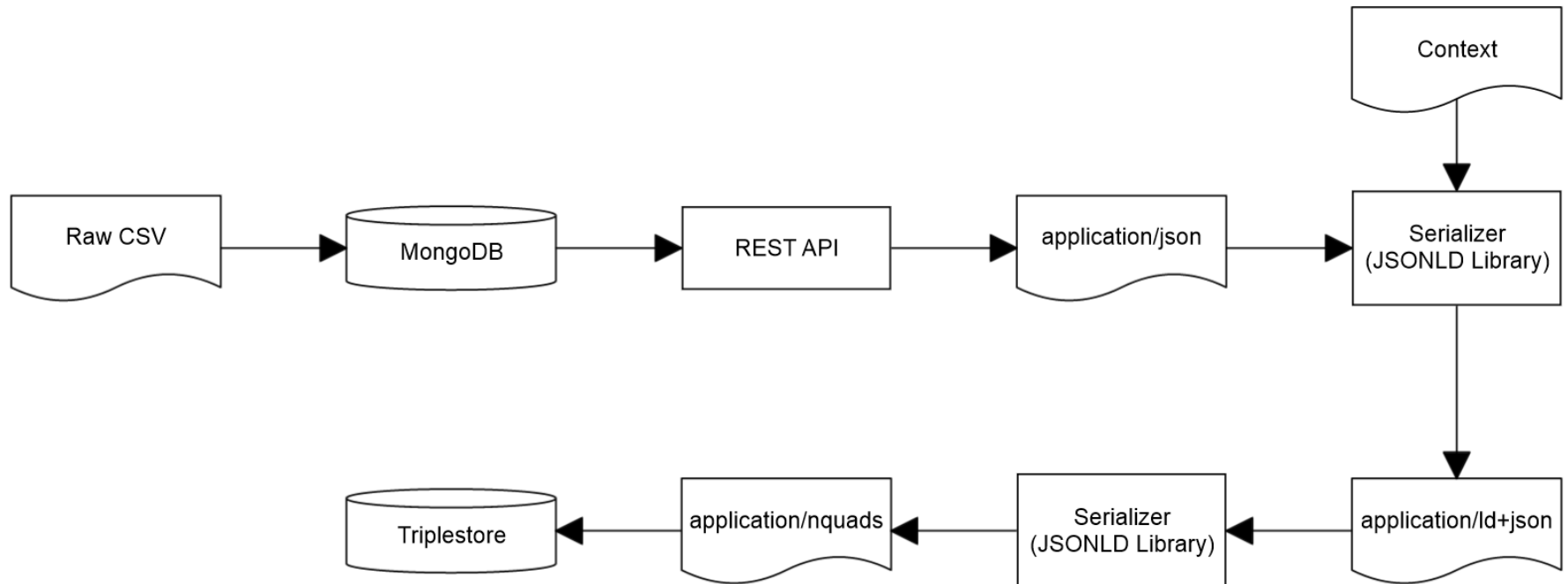
1: JSON support



2: JSON-LD support



3: Triples support



Gebruikte technologie

- AngularJS frontend
- RAML modeling language
- MongoDB raw data store
- AllegroGraph triplestore
- Symfony2 PHP API



ANGULARJS

RAML



Symfony



mongoDB



AllegroGraph

Franz Inc.



freshheads
nieuw internet

“Linked Data maakt het mogelijk om voorheen technisch onmogelijke inzichten te krijgen in gegevens. Het is nu aan de marketeers om interessante vraagstukken te verzinnen met deze nieuwe schat aan informatie.”

Vragen?

- Twitter: <http://twitter.com/dvh>
- E-mail: dimitri@freshheads.com
- GitHub: <http://github.com/dvh>
- LinkedIn: <https://linkedin.com/in/dimitrivanhees>